Intensive Care Med (2010) 36:1207–1212
DOI 10.1007/s00134-010-1852-2

ORIGINAL

Rui P. Moreno
Helene Hochrieser
Barbara Metnitz
Peter Bauer
Philipp G. H. Metnitz

Characterizing the risk profiles of intensive care units

Received: 12 October 2009
Accepted: 13 January 2010
Published online: 20 March 2010

On behalf of the ASDI Study Group.

Electronic supplementary material
The online version of this article (doi:10.1007/s00134-010-1852-2) contains supplementary material, which is available to authorized users.

R. P. Moreno
Unidade de Cuidados Intensivos Polivalente, Hospital de Santo António dos Capuchos, Centro Hospitalar de Lisboa Central, E.P.E., Lisbon, Portugal

H. Hochrieser · B. Metnitz · P. Bauer (✉)
Section for Medical Statistics, Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria
e-mail: peter.bauer@meduniwien.ac.at

P. G. H. Metnitz
Department of Anesthesiology and General Intensive Care, Medical University of Vienna, Vienna, Austria

Abstract *Objective:* To develop a new method to evaluate the performance of individual ICUs through the calculation and visualisation of risk profiles. *Methods:* The study included 102,561 patients consecutively admitted to 77 ICUs in Austria. We customized the function which predicts hospital mortality (using SAPS II) for each ICU. We then compared the risks of hospital mortality resulting from this function with the risks which would be obtained using the original function. The derived risk ratio was then plotted together with point-wise confidence intervals in order to visualise the individual risk profile of each ICU over the whole spectrum of expected hospital mortality. *Main measurements and results:* We calculated risk profiles for all ICUs in the ASDI data set according to the proposed method. We show examples how the clinical performance of ICUs may depend on the severity of illness of their patients. Both the distribution of the Hosmer–Lemeshow goodness-of-fit test statistics and the histogram of the corresponding *P* values demonstrated a good fit of the individual risk models. *Conclusions:* Our risk profile model makes it possible to evaluate ICUs on the basis of the specific risk for patients to die compared to a reference sample over the whole spectrum of hospital mortality. Thus, ICUs at different levels of severity of illness can be directly compared, giving a clear advantage over the use of the conventional single point estimate of the overall observed-to-expected mortality ratio.

Keywords Risk adjustment · Risk stratification · Outcome · Intensive care · Severity of illness

Introduction

Finding a reliable method to quantify the performance of single ICUs has always been a difficult quest in the last 30 years, despite the fact that the evaluation of clinical performance is a prerequisite for the assessment of both the effectiveness and efficiency of care. Furthermore, increasing cost constraints have focused the attention of clinicians, managers and researchers on this problem.

Most published approaches to quantifying the performance of ICUs adopt more or less the same procedures: the development of a general outcome prediction model (GOPM) and its calibration in a suitable database. Such models are then applied to different cohorts of ICU patients, and comparison of the predicted number of deaths with the actual number of deaths is used as a reference for the clinical performance of the unit. The resulting quotient (actual number of deaths divided by

predicted number of deaths) is thus a one-point estimate, known as the standardized mortality ratio (SMR) or observed-to-expected mortality ratio (*O*/*E*). If this ratio is significantly lower than 1, then the performance of the specific ICU being evaluated is judged to be better than that of the ICUs from which the GOPM was derived; if it is significantly higher than 1, the performance of the ICU is judged to be worse.

Using several GOPMs, such as APACHE II [1], APACHE III [2], APACHE IV [3], SAPS II [4] or SAPS 3 [5], this methodology has become the "gold standard" to compare ICUs across different geographical areas or inside the same country, or other specific subgroups of patients. Several limitations of the SMR approach, including patient-, user-, and model-dependent problems, have been described and discussed in detail [6]. But there is another weakness: a one-point estimate considers the performance of an ICU to be constant over the whole spectrum of the severity of illness. In other words, an ICU with a "good" performance (low SMR) is assumed to be uniformly good for both low-risk and high-risk patients; likewise, an ICU with a "bad" performance (high SMR) is assumed to be uniformly bad. However, this assumption is most probably not true, since performance might differ over the spectrum of the severity of illness.

The objective of this work was to develop a new method to evaluate the clinical performance of ICUs, on the assumption that clinical performance might be variable over the spectrum of severity of illness of the admitted patients. In this report we present several examples of risk profiles for ICUs using data aggregated between 1998 and 2007.

Methods

The study used the database of the Austrian Center of Documentation and Quality Assurance in Intensive Care Medicine (ASDI), a nonprofit organization that has established an intensive care database and benchmarking project in Austria. The prospectively collected data included sociodemographic data, such as age, sex and chronic conditions; reason for admission, which was recorded according to a predefined list of medical and surgical diagnoses; severity of illness, as measured by SAPS II [4]; length of ICU stay and hospital stay; and outcome data, including survival status at ICU and hospital discharge.

Data on all patients consecutively admitted to 77 Austrian ICUs between January 1, 1998, and December 31, 2007, were included in this study. A total of 176,703 admissions occurred during the study period. For readmitted patients, only the first admission was included, leaving 168,194 patients. Patients who were <18 years of age ($n = 3{,}292$), those with records that lacked an entry in the field "hospital outcome" ($n = 6{,}904$) and those without a valid SAPS II score ($n = 1{,}313$) were further excluded.

To avoid over- or underestimating the relative risks for patients in the extremely low and extremely high risk areas, we also excluded patients with a predicted hospital mortality of <5% ($n = 52{,}181$) or >95% ($n = 1{,}937$). Predicted mortality was based on the original SAPS II risk model. Consequently, the cohort under study consisted of 102,561 patients from 77 ICUs. Since no additional interventions were performed, the need for informed consent was waived by the institutional review board.

Using SAPS II [4], we customized the function which predicts hospital mortality at ICU admission for each ICU. We used simple customization in complete samples. The risks of hospital mortality resulting from this function were then compared with the risks which would be obtained using a reference cohort. As reference we used the original function from the SAPS II publication applied to the ASDI cohort. By dividing the risk obtained from the ICU-specific function by the risk obtained from the general function, we were able to determine the difference in the risk of dying in the hospital for patients in individual ICUs over the continuum of 5–95% predicted risk. The logistic regression-derived risk ratio was then graphically plotted, together with its confidence intervals. Conventional point-wise 95% confidence intervals for the risk ratios were calculated from the customized logistic models assuming that the expected risk from the original SAPS II risk model in the denominator is a non-random quantity.

The result is a risk profile of each ICU over the whole span of the probability of hospital mortality. This method thus allows a direct comparison of individual risk profiles for specific ICUs. Using a simple two parameter risk model with a form closely related to the model used originally when constructing the SAPS II risk score achieves smoothing of the individual risk profiles with corresponding confidence intervals in moderately small samples. This is thus different to calculating ICU-specific risk ratio estimates in risk decile groups with even much smaller sample sizes (and potentially wide confidence intervals).

To evaluate the performance of these risk profiles, we compared their results with those estimated by nonparametric methods through the use of LOWESS curves (SAS, Proc LOWESS). We used three different methods for this purpose: first, a graphical distribution of the Hosmer–Lemeshow goodness-of-fit test per ICU per year, which (for a perfect fit) approximately should follow a chi-square distribution with 8 degrees of freedom (*df*). Second, a histogram of the corresponding *P* values per ICU per year. Third, a display of the corresponding LOWESS curves together with the regression-based

estimates for each ICU to allow for visualization of the results.

Results

A total of 102,561 patients were included in the study. The 77 ICUs provided a median number 1,135 (interquartile range 512–1,905) patients per ICU. Median age of the patients was 70 (60–79) and 44% were female. Roughly half of the admissions were medical admissions (Table 1). Chronic cardiac failure was the leading comorbidity (17.8%), followed by chronic renal failure (8.9%), metastatic cancer (6.1%) and insulin-dependent diabetes mellitus (6.1%).

Figure 1 shows (as an explanatory example) the association between the SAPS II score and the predicted hospital mortality for the reference cohort and for a specific ICU (the same ICU as in Fig. 2a). The risk of dying in the hospital for a patient with a SAPS II score of approximately 29 was 10% in the reference cohort, but roughly 17% in the individual ICU. Dividing the risk in the ICU by the risk in the reference cohort, we obtained the risk ratio as depicted in Fig. 2a: in this case, 1.7 for the risk point of 0.1 on the *x*-axis, which corresponds to the increased risk of dying as outlined above. This risk ratio was calculated for the whole span of the severity of illness.

Calculating the risk profiles for all ICUs from the ASDI database we found several units whose performance depended on the severity of illness of its patients. Figure 2 gives some examples of different relationships between the actual risk of dying in a specific ICU and the risk predicted in the reference cohort. Figure 2a shows data for an ICU in which clinical performance increased as the severity of illness increased, i.e., fewer patients died than expected. Thus, this ICU performed better with high-risk patients. The opposite was true for the example shown in Fig. 2b, in which clinical performance decreased as the severity of illness increased; more patients died than expected in the upper risk area. Thus, this ICU performed better with low-risk patients. Figure 2c shows the example of an ICU in which clinical performance was stable over the whole spectrum of the severity of illness.

Figure 2 shows the results for three individual ICUs aggregated over a time period up to 10 years. In order to get an impression of the variability of the risk profiles in smaller samples Figure E1 in the electronic material provides the corresponding results of the same three ICUs broken down to calendar years (altogether 504 samples). The most interesting result is that the different types of risk profiles for specific ICUs seem to remain surprisingly stable over the years in the ICUs.

Comparing the risk profiles estimated by our risk profile model with those estimated by nonparametric methods (LOWESS curves), we found a satisfactory fit of the risk profile model: results from the simple two parameter risk model in general are rather closely following the profile resulting from nonparametric smoothing of an arbitrary type of profile (Fig. 2; Figure E1). The histograms for the Hosmer–Lemeshow goodness-of-fit test statistics per ICU per year showed excellent agreement with the density of the chi-square distribution with 8 *df* (Fig. 3a). The histogram of the

Table 1 Characteristics of the study subjects

		Percentage
Number of patients	102,561	100
Number of ICUs	77	100
Patients per ICU	1,135 (512–1,905)	
Age (years)	71.0 (60.0–79.0)	
Sex		
Female	44,738	43.74
Male	57,540	56.26
Missing	283	
ICU length of stay (days)	4.00 (2.00–9.00)	
Total number of days in the ICU	873,776	
Outcome		
SAPS II score	36.0 (28.0–49.0)	
SAPS II predicted mortality	29,484	28.75
Observed ICU mortality	16,801	16.38
Observed hospital mortality	24,642	24.03
Type of admission		
Medical	52,379	51.22
Scheduled surgical	24,570	24.03
Unscheduled surgical	25,307	24.75
Missing	305	
SAPS II *O*/*E* ratio	0.836 (0.828–0.844)	

All data are given as median with interquartile ranges if not otherwise expressed

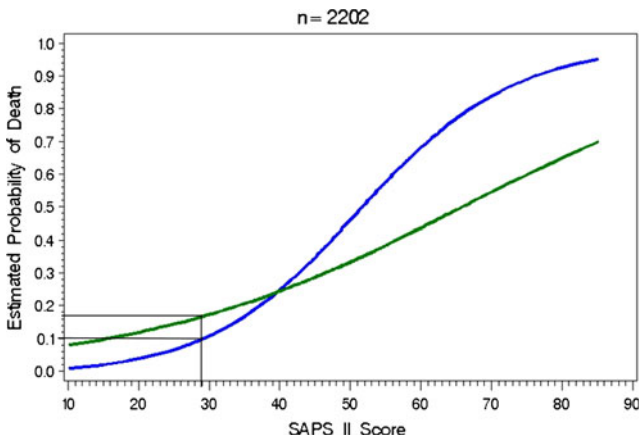

Fig. 1 Association between SAPS II score and predicted mortality in the reference cohort and a single ICU. The *blue line* represents the relationship between the SAPS II score and the predicted probability of death using the original SAPS II function. The *green line* represents the customized function for the specific ICU

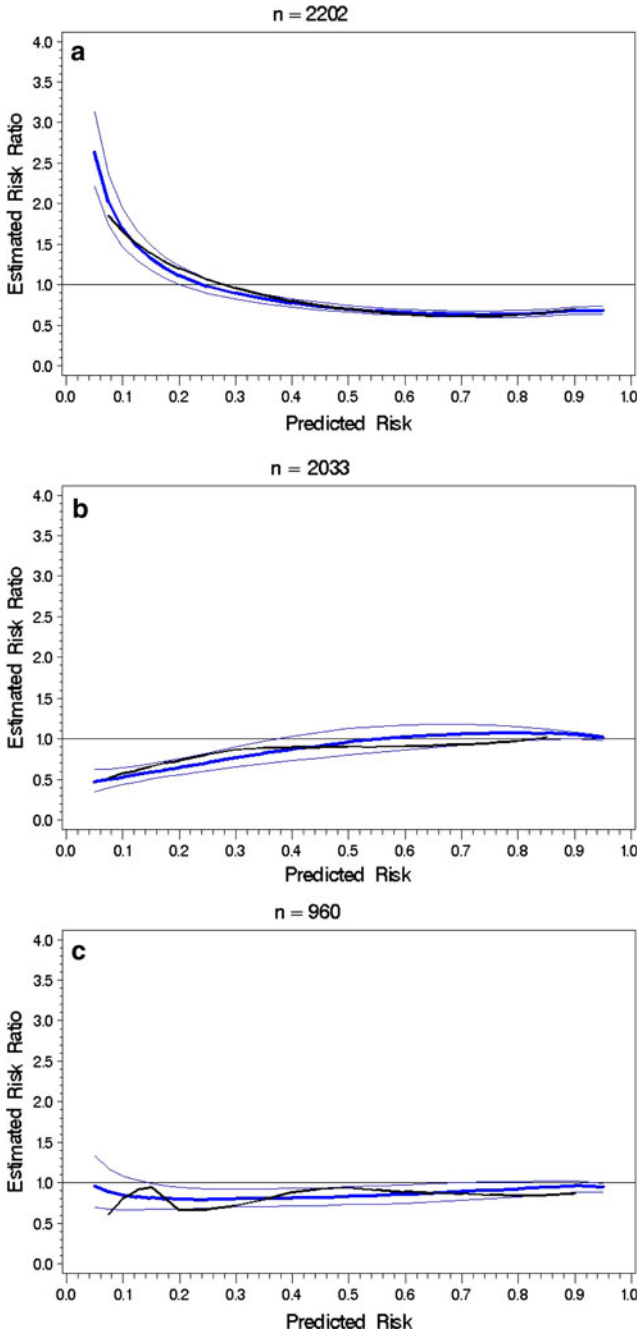

Fig. 2 Examples of different relationships between the performance of the ICU and the severity of illness of the admitted patients. The *thick blue lines* represent the risk profile model. The *thinner blue lines* represent the confidence interval. The *black lines* represent the associated LOWESS curve. Data were aggregated between 1998 and 2007. **a** ICU in which the clinical performance increased as the severity of illness increased. **b** ICU in which the clinical performance decreased as the severity of illness increased. **c** ICU in which the clinical performance was almost independent of the severity of illness

corresponding *P* values was also plotted (Fig. 3b). For a perfect fit in independent samples the distribution should be uniform between 0 and 1. As can be seen in Fig. 3b, there was no obvious tendency toward low *P* values (which would be expected for a lack of fit), and thus we can conclude that the risk profile model again demonstrates a sufficient fit.

To see how the model works if all patients are included in the analysis (including those below 5 and above 95% risk) we repeated the calculations for the whole sample ($n = 156.679$) (with arbitrary expected risk) fulfilling the inclusion criteria. Figures E2, E3 and E4 in the electronic material, give the corresponding results for the same ICUs as in Fig. 2a–c in this manuscript, respectively. Overall the fit in the full samples tends to be a little worse at the extreme ends of the risk scale (as could be awaited), but the confidence intervals are narrower due to the larger sample sizes particularly in the low risk area.

Discussion

Unlike the one-point estimate used in the classical SMR approach, the model presented here provides a graphic representation of the risk profile for any ICU over the whole spectrum of predicted hospital mortality. Contrary to the classical view, in which clinical performance is assumed to be the same in all risk groups, we tested and accepted the hypothesis that performance can vary not only between ICUs, but also within the same ICU, according to the severity of illness of the admitted patients.

As shown in Fig. 2, variations in clinical performance do exist. In an attempt to address this problem, investigators analyzed the EURICUS-I study database [7] by deriving three performance variables for each ICU: P20 (performance with low-risk patients, evaluated at the 20th percentile), P50 (performance with medium-risk patients, evaluated at the median) and P80 (performance with high-risk patients, evaluated at the 80th percentile). This method, based on customization of the SAPS II score for each ICU using logistic regression with random effects, allowed the user to compare the performance of an ICU at least three different points. However, this method was never widely used. Similar results were obtained during the development of the SAPS 3, when the relationship between severity of illness and actual mortality was found to differ greatly between different geographical areas [6].

The conceptual and practical limitations of the concept that clinical performance is independent of severity of illness cannot be overemphasized, because the implication (used by all current GOPMs) is, that performance is

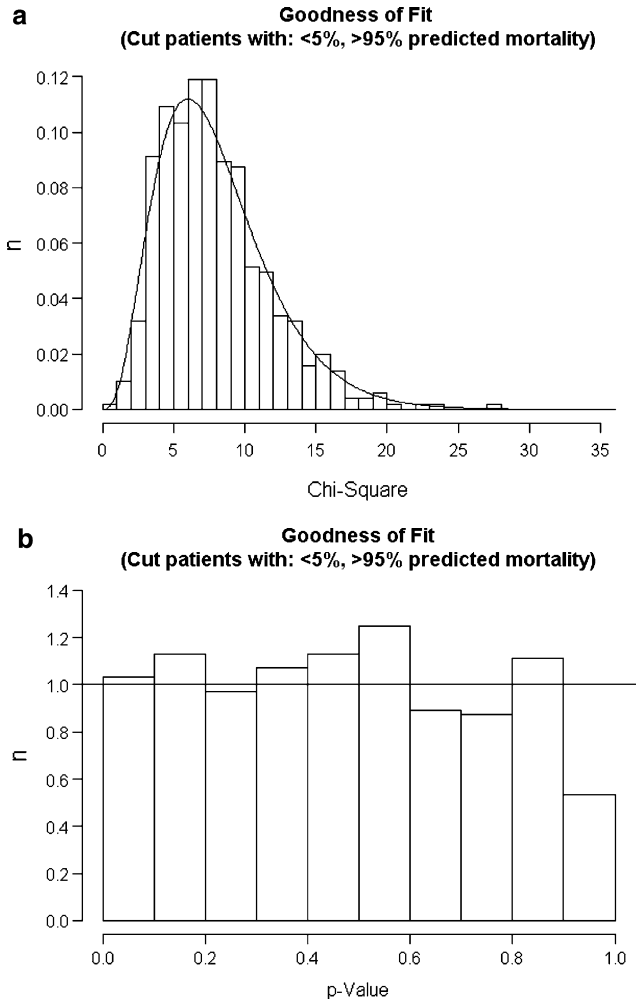

Fig. 3 Evaluation of the risk profile model. **a** Histogram of the Hosmer–Lemeshow goodness-of-fit test statistics per ICU per year (in total 504 samples) from the SAPS II analyses. The *curve* indicates the density of the chi-square distribution with 8 *df*, the distribution to be expected approximately for the case of a perfect fit in independent samples. **b** Histogram of the corresponding *P* values, which for a perfect fit in independent samples should approximately follow a uniform distribution between 0 and 1. Also, there is no obvious tendency toward low *P* values (lack of fit)

stable—that ICUs always perform badly or well. However, as we demonstrate here, performance may even vary within ICUs. This is true not only for variations over time [8], but also over the spectrum of severity of illness. This means that the same ICU can present a good performance when dealing with patients with low severity of illness and a poor performance with patients with high severity of illness, or vice versa. When looking at the variation of such profiles over calendar years, we saw examples of ICUs with surprisingly stable profiles (Figures E1, E4). The challenge is thus to first understand why this phenomenon occurs and then to determine which factors (clinical or nonclinical) are responsible for these variations in performance.

Most ICUs have been designed to deal with a certain type of patient and a certain number of patients and thus are equipped with the technical and human resources strictly necessary for this task. What happens, however, if this capacity is overextended? There is good evidence that several operational factors affect the outcome of critically ill patients, such as high nursing workload [9], increased occupancy rate [10], and patient turnover [11]. However, the opposite could also be true: an ICU that is used to deal with very critically ill patients could have a worse performance when only low-risk patients are admitted. This could be explained in part by the overmonitoring and overtreating of patients, which has been shown to affect outcome as well [12, 13].

For purposes of risk adjustment, we used the SAPS II system in this study. However, it should be emphasized that any risk prediction model which allows for a transformation of the score into a predicted risk of death could be used. We used SAPS II because it is available—it is scored on a mandatory basis in Austria. We also tested the risk profile model in the SAPS 3 database and found good results there as well (data not shown). However, the number of observations in the ASDI database is much higher because of the long period of data collection, and thus we decided to present those data here. Anyway, further studies using other risk adjustment systems should be undertaken.

Moreover, we used the original SAPS II equation as a reference in this study. This could be questioned, since it is well known that the original equation overestimates hospital mortality in Austrian ICU patients [the *O*/*E* ratio for the original SAPS II from the same cohort is 0.841 (0.834–0.849)]. Since we do not compare single *O*/*E* mortality ratios, this deviation does not really distort our results. For future studies, however, the reference database (and thus the reference function for the prediction of hospital mortality) could be chosen according to the purpose of comparison: if necessary, regionally customized functions (when available, e.g., in the SAPS 3 system) could be used instead or in addition.

We excluded patients within the extremely low and extremely high risk areas, i.e., those with a predicted mortality of <5 or >95%. Although the number of patients excluded seems to make up a substantial proportion of the whole cohort, the numbers diminish if the relatively high number of ICUs is taken into account. The exclusion seems to be advisable, since low numbers of patients with extreme risks would affect the values of the expected risk in the denominator and thus distort the results of the model. For this reason we decided first to develop the model applying this restriction.

However, even in the unrestricted patient population, the model produces very reasonable profiles with moderately narrower confidence intervals (due to the

increased sample sizes). In future it should be the quality of prediction in extreme risk categories of the underlying risk model which should taken as a basis for the decision, if at all and how any truncation should be performed.

In conclusion, our results, derived from a large cohort of critically ill patients, provide conclusive evidence that the clinical performance of ICUs may vary over the spectrum of severity of illness. These results provide the rationale to further investigate the association between these variations happening and organizational factors, so as to detect possibilities to improve the care we deliver to our patients.

Acknowledgments The study was supported by the Austrian Centre for Documentation and Quality Assurance in Intensive Care Medicine (ASDI). Our special thanks to the participants from all the ICUs that participated in the project.

References

1. Knaus WA, Draper EA, Wagner DP, Zimmerman JE (1985) APACHE II: a severity of disease classification system. Crit Care Med 13:818–829
2. Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, Sirio CA, Murphy DJ, Lotring T, Damiano A, Harrell FE Jr (1991) The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. Chest 100:1619–1636
3. Zimmerman JE, Kramer AA, McNair DS, Malila FM (2006) Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. Crit Care Med 34:1297–1310
4. Le Gall JR, Lemeshow S, Saulnier F (1993) A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. JAMA 270:2957–2963
5. Moreno RP, Metnitz PG, Almeida E, Jordan B, Bauer P, Campos RA, Iapichino G, Edbrooke D, Capuzzo M, Le Gall JR, SAPS 3 Investigators (2005) SAPS 3. From evaluation of the patient to evaluation of the intensive care unit. Part 2: development of a prognostic model for hospital mortality at ICU admission. Intensive Care Med 31:1345–1355
6. Metnitz PG, Moreno RP, Almeida E, Jordan B, Bauer P, Campos RA, Iapichino G, Edbrooke D, Capuzzo M, Le Gall JR, SAPS 3 Investigators (2005) SAPS 3. From evaluation of the patient to evaluation of the intensive care unit. Part 1: objectives, methods and cohort description. Intensive Care Med 31:1336–1344
7. Reis Miranda D, Ryan DW, Schaufeli WB, Fidler V (eds) (1997) Organization and management of intensive care: a prospective study in 12 European countries. Springer, Berlin
8. Lovegrove J, Valencia O, Treasure T, Sherlaw-Johnson C, Gallivan S (1997) Monitoring the results of cardiac surgery by variable life-adjusted display. Lancet 350:1128–1130
9. Moreno R, Reis Miranda D (1998) Nursing staff in intensive care in Europe. The mismatch between planning and practice. Chest 113:752–758
10. The UK Neonatal Staffing Study Group (2002) Patient volume, staffing, and workload in relation to risk adjusted outcomes in a random stratified sample of UK neonatal intensive care units: a prospective evaluation. Lancet 359:99–107
11. Metnitz B, Metnitz PG, Bauer P, Valentin A, ASDI Study Group (2009) Patient volume affects outcome in critically ill patients. Wien Klin Wochenschr 121:34–40
12. Metnitz PG, Reiter A, Jordan B, Lang T (2004) More interventions do not necessarily improve outcome in critically ill patients. Intensive Care Med 30:1586–1593
13. Connors AF, Speroff T, Dawson NV, Thomas C, Harrell FE Jr, Wagner D, Desbiens N, Goldman L, Wu AW, Califf RM, Fulkerson WJ, Vidaillet H, Broste S, Bellamy P, Lynn J, Knaus WA, The SUPPORT investigators (1996) The effectiveness of right heart catheterization in the initial care of critically ill patients. JAMA 276:889–897